

Web Page:	https://studyinmexico.tec.mx/
Contact Information:	studyinmexico@itesm.mx

Undergraduate Research Program	
Project Name	Protein 3D structure prediction using compression formats based on musical nomenclature
Campus & Location in Mexico	Monterrey
Faculty	Engineering
Research Area	Computational Learning Models
Research Responsible	Dr. Edgar Acuña González
Description of the Project	Great algorithms, such as AlphaFold2, have already found a way to predict with great accuracy 3D structures of proteins only using a single chain of aminoacids as input. These algorithms have a great impact in biotechnology because they save a lot of resources in predicting the function of a protein, since similar motives within the structure relate to similar functions of proteins. However, there are certain proteins - such as isoenzymes - that have the same function but an apparently different 3D protein structure which makes it hard to predict why exactly these proteins have similar function. Thus, the next step is to improve the algorithm by adding more information into the machine-learning method so new patterns can be discovered that explain protein functions. By simply adding more information, however, the issue won't be solved because it would require exponentially more data processing hence making the algorithm useless. One way to solve this issue is to compress the information in a way that with few data points and specifying their relationship within you can store more information in the same data space, similar to what happened with the appearance of Blue-ray vs. DVD. A well-known method for compressing information that has not been so explored in these areas is the musical score where musicians are able to deduce musical structures, functions and relationships with just 7 bits of information (musical notes). So, the objective of this project is to utilize musical score concepts to get as much information as possible using few data points and relate them to the most relevant aspects of protein structures to run an unsupervised machine-learning method and find new patterns of protein functions that apparently are not related to similarities between aminoacid chains.
Training Provided	Design of experiments;Prototype development;Test hypothesis
Modality	Hybrid

Offered During	Summer (5 weeks);Winter (5 weeks);Semester
-----------------------	--

Student	
Tasks/Responsibilities	Read relevant scientific material, contribute to hypothesis generation, annotate detailed failed/successful experiments, discuss of further steps to follow to accomplish project's objectives.
Required Language Proficiency	English (Advanced)
Required Skills and Abilities	Python and machine learning algorithms (medium to advance)
Other Documents Required to APPLY for an Internship	1) Being at least in your 2nd year of bachelor 2) Accumulative grade point average (GPA) 2.5 3) Official Transcript 4) 2 letters of recommendation of faculty members 5) Resume 6) Letter of intention explaining the reason why you would like to participate in the research program